

Big Data Praxisbeispiel – Cloud-Computing

Viele individuelle Daten werden mittlerweile über Dienste wie Dropbox oder Google Drive in einer sogenannten Cloud gespeichert, z. B. weil die Speicher- und Rechenkapazitäten des eigenen Computers nicht ausreichen, um mit den vorhandenen Datenmengen umzugehen. Das „Cloud Computing“ bezeichnet demnach Prozesse, bei denen Daten an externen, zentralen Speicherorten (in der „Cloud“) abgelegt und genutzt werden.

Der nachfolgende Text erläutert, wie Cloud-Computing funktioniert, wie es sinnvoll eingesetzt wird und wie Online-Übersetzungsdienste sich der dadurch entstehenden Möglichkeiten bedienen.

Was ist Cloud-Computing genau?

Übersetzt ins Deutsche bezeichnet Cloud-Computing das Arbeiten mit oder in einer Rechner- oder Datenwolke. Über das Internet stehen Privatpersonen und Unternehmen leistungsfähige Speicher- oder Rechendienste in entfernten Rechenzentren zur Verfügung. Dafür wählen sich Computer oder Smartphones zum Beispiel über einen Webbrowser oder Apps in riesige, zentrale Server- und IT-Infrastrukturen ein. Auch Funktionen des Betriebssystems sorgen dafür, dass sich die Geräte im Hintergrund mit der Cloud verbinden, beispielsweise um Zugänge zu WLANs oder E-Mail-Konten zentral abzulegen. So ist es möglich, von mehreren Geräten aus und weltweit auf sie zuzugreifen. Voraussetzung dafür ist, dass man entsprechende Zugangsdaten eingibt oder hinterlegt. Zu den bekanntesten, weltweit millionenfach genutzten Clouds gehören Dienste von Google, Amazon, Apple, Microsoft oder Dropbox.

Cloud-Computing-Dienste kommen sowohl in der Wirtschaft und der Wissenschaft als auch im privaten Bereich zum Einsatz. Generell unterscheidet man drei Cloud-Computing-Bereiche: Die strukturierte Datenspeicherung, das Bereitstellen von Rechenleistung sowie das Bereitstellen von Anwendungsprogrammen. Die Software dafür ist in der Cloud installiert und nicht auf dem eigenen PC, Tablet oder Smartphone. Mit ihrer Hilfe ist es möglich, Texte zu bearbeiten, Präsentationen zu erstellen, mathematische Berechnungen durchzuführen oder eben Sprachübersetzungen zu nutzen.

Was hat das mit Big Data zu tun?

Generell bietet sich Cloud-Computing immer dann an, wenn viele Nutzerinnen und Nutzer an zahlreichen unterschiedlichen Orten möglichst einfach und gleichzeitig Zugriff auf dieselben Daten und Services haben sollen. Oder wenn die zu verarbeitenden Datenmassen so groß und die Anforderungen an die Programme so hoch sind, dass weder heimische Computer noch Firmenrechner, Smartphones oder Tablets die Aufgaben bewältigen können.

Am Beispiel cloudbasierter Übersetzungsdienste, zu denen unter anderem der Google Übersetzer oder DeepL zählen, zeigt sich, welche Rolle Big Data hierbei spielen kann.

Die cloudbasierten Sprachübersetzungsdienste greifen auf gigantische Datensammlungen von bereits übersetzten Texten zurück. Google setzt seit einigen Jahren darauf, dass tausende Nutzerinnen und Nutzer Übersetzungen hochladen bzw. eingeben oder bearbeiten. Die Entwicklerinnen und Entwickler von DeepL hingegen programmierten sogenannte „Crawler“, die das Internet weltweit nach fertig übersetzten Texten durchsuchen. Mit dem systematisch anwachsenden Datenbestand bauten sie dann eine spezielle Suchmaschine nur für Übersetzungen. Diese wurde immer häufiger genutzt, was wiederum dazu führte, dass die Betreiber noch mehr Übersetzungen in die Big-Data-Bestände übernehmen konnten.

Die kontinuierlich gewachsenen Datenbanken mit präzise übersetzten Texten werden nun mit Big-Data-Analysen ausgewertet, um möglichst gelungene Übersetzungen zu erhalten, die sprachlich

verständlich und inhaltlich korrekt sind. Dafür werden Wortfolgen, Satzstellungen und Redewendungen praktisch wie Muster betrachtet, die sich wiederholen und häufen. Sie werden dann mit Wortfolgen, Satzstellungen und Redewendungen der anderen Sprache verglichen. Je öfter sich die jeweiligen Muster der einen Sprache mit denen der anderen decken, desto wahrscheinlicher ist diese Übersetzung korrekt. Diese Vorgänge laufen mithilfe von Algorithmen und künstlicher Intelligenz automatisiert ab. Damit dieser Dienst für die Nutzerinnen und Nutzer in Echtzeit funktioniert – online und mitunter tausendfach gleichzeitig – sind sowohl große Datenbanken als auch enorme Rechenleistungen notwendig, wie sie Supercomputer und sogenannte Serverfarmen bereitstellen.

Was ist daran bedenklich?

Mit Cloud-Computing ist es demnach möglich, die eigenen Daten auszulagern, um nicht an die Grenzen der eigenen Hardware zu geraten. Darüber hinaus ergeben sich zugleich neue Möglichkeiten, massenhaft Daten zu kombinieren und damit im Ergebnis gute und im Prozess schnelle Dienstleistungen anzubieten, die den Nutzerinnen und Nutzern zur Verfügung stehen. Übersetzungsdienste, die auf riesige Bestände übersetzter Texte zugreifen, die in der Cloud gespeichert sind, sind für sich genommen erst einmal nicht bedenklich, weil sie keine personenbezogenen Daten für ihre Zwecke auswerten oder neu kombinieren. Generell neigen Nutzerinnen und Nutzer durch die Verfügbarkeit von Cloud-Diensten dazu, immer mehr Daten mit persönlichen Informationen in der Cloud zu speichern und nicht auf der eigenen Festplatte. Auf diese Weise wird es für andere Akteure leichter, auf personenbezogene Daten zuzugreifen und sie für Big-Data-Anwendungen auszuwerten. Formuliertes Ziel dabei ist, die eigenen Angebote noch besser zu machen. Tatsächlich geht es häufig jedoch darum, das eigene Geschäftsmodell zu optimieren, z. B. auch für mögliche Werbekunden, oder darum, die eigene Marktmacht zu festigen und auszubauen.

Mit dem Speichern der Daten in der Cloud begibt sich jede und jeder Einzelne in die Abhängigkeit eines Dienstleisters, von dem oft wenig bekannt ist. So kam es in der Vergangenheit immer wieder zu stundenlangen Ausfällen von Cloud-Diensten und die dort gespeicherten Daten waren in der Zeit nicht verfügbar. Manche Dienste wurden vom Anbieter ganz eingestellt, sodass die Daten oft mühsam zu einem anderen Anbieter übertragen werden mussten. Und nicht zuletzt wurden auch schon persönliche Daten durch Hackerangriffe von den Cloud-Servern gestohlen.